Article

Matlab methods for calculation of between-image correlations and similarities

WenJun Zhang[1], **Yanhong Qi**[2]
[1]School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China; International Academy of Ecology and Environmental Sciences, Hong Kong
[2]Libraries of Sun Yat-sen University, Sun Yat-sen University, Guangzhou, China
E-mail: zhwj@mail.sysu.edu.cn, wjzhang@iaees.org

Received 1 November 2019; *Accepted 16 November 2019*; *Published 1 December 2019*

Abstract
In present study, we developed a series of methods to calculate between-image correlations and similarities. Between-image Pearson correlation, Spearman correlation and similarity index can be achieved and various statistic tests for between-image difference, including signed test of matched samples for between-image difference, Wilcoxon signed rank test for zero median of between-image difference, and *t*-test for between-image difference, can be made by using these methods. Full Matlab codes were given.

Keywords Matlab; methods; image analysis; correlation; similarity; Wilcoxon test.

Ornamental and Medicinal Plants
ISSN 2522-3682
URL: http://www.iaees.org/publications/journals/omp/online-version.asp
RSS: http://www.iaees.org/publications/journals/omp/rss.xml
E-mail: omp@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Image analysis is one of the core contents in plant and medical studies and applications. There are many methods for image analysis, e.g., image denoising, image adjustment, edge detection, etc. However, we are short of methods for between-image comparion and alignment. In this study, we will present some methods to calculate between-image correlations and similarities. Between-image Pearson correlation, Spearman correlation and similarity index can be achieved and various statistic tests for between-image difference can be made by using these methods. Full Matlab codes will be given.

2 Materials and Methods

Suppose we compare two gray images, represented by two matrices $A=(a_{ij})$ and $B=(b_{ij})$, $i=1, 2, ..., m$; $j=1, 2, ..., n$, where m and n are the number of pixels along y-axis and x-axis, $0\leq a_{ij}\leq 1$ and $0\leq b_{ij}\leq 1$ are the standardized gray values at pixel (i, j) of two images respectively. First, two images are transformed to gray images. Resize the two images to the same dimension of $m\times n$, where m and n are the means of dimensions of two images respectively. The gray values of images are standardized as values in [0,1].

2.1 Pearson correlation and Spearman correlation between two images

Matrices A and B are transformed to vectors A_1 and B_1 respectively, i.e., $(A_1, B_1)=\{(a_i, b_i) \mid i=1, 2, ..., p\}$ (Zhang, 2018b). By using Pearson correlation and Spearman correlation (Zhang, 2012, 2015, 2018a-b; Zhang and Li, 2015), we can achieve the test results for between-image correlation. A greater Pearson correlation and Spearman correlation means the higher similarity in image structure between two images.

Full Matlab codes for Pearson correlation and Spearman correlation between two images are as follows (Zhang, 2012, 2015, 2018a-b):

```
sig=input('Input significance level(e.g., 0.01): ');
RGB1=imread('geneExpression.png'); RGB2=imread('soluableProteins.png');
Gray1=rgb2gray(RGB1); Gray2=rgb2gray(RGB2);
m=round((size(Gray1,1)+size(Gray2,1))/2); n=round((size(Gray1,2)+size(Gray2,2))/2);
I=imresize(Gray1,[m n]); J=imresize(Gray2,[m n]);
A=im2double(I); B=im2double(J);
A1=reshape(A, m*n, 1);    %Transform matrix A to vector A1
B1=reshape(B, m*n, 1);    %Transform matrix B to vector B1
r=corr(A1,B1);
rsp=spearman(A1,B1);
fprintf(['Pearson correlation r=' num2str(r) '\n']);
fprintf(['Spearman correlation r=' num2str(rsp) '\n']);
tvalue=abs(r)/sqrt((1-r^2)/(m*n-2));
tspvalue=abs(rsp)/sqrt((1-rsp^2)/(m*n-2));
p=(1-tcdf(tvalue,m*n-2))*2;
psp=(1-tcdf(tspvalue,m*n-2))*2;
sigma=p<sig;
sigmasp=psp<sig;
if (sigma==1) fprintf(['Pearson correlation is statistically significant (p=' num2str(p) ')\n']); end
if (sigma==0) fprintf(['Pearson correlation is not statistically significant (p=' num2str(p) ')\n']); end
if (sigmasp==1) fprintf(['Spearman correlation is statistically significant (p=' num2str(psp) ')\n']); end
if (sigmasp==0) fprintf(['Spearman correlation is not statistically significant (p=' num2str(psp) ')\n']); end

function spearm =spearman(x,y)             %x and y: two column vectors to be tested.
if (max(size(x))~=max(size(y)))
error('Array sizes do not match.');
end
if ((min(size(x))~=1) | (min(size(y))~=1))
error('Both x and y are vectors');
end
n=max(size(x));
for i=1:n
rx(i)=0;ry(i)=0;xx(i)=0;yy(i)=0;
end
for j=1:n
nx=1;ny=1;
for i=1:n
```

```
if (x(i)<x(j)) nx=nx+1; end
if (y(i)<y(j)) ny=ny+1; end
end
rx(j)=nx;
ry(j)=ny;
end
for j=1:n
if (rx(j)==(n+1)) continue; end
nx=rx(j);
ntie=-1;
for i=1:n
if (rx(i)~=nx) continue; end
ntie=ntie+1;
xx(i)=rx(i);
rx(i)=0;
end
for i=1:n
if (rx(i)~=0) continue; end
xx(i)=xx(i)+(ntie*0.5);
rx(i)=n+1;
end
end
for j=1:n
if (ry(j)==(n+1)) continue; end
ny=ry(j);
ntie=-1;
for i=1:n
if (ry(i)~=ny) continue; end
ntie=ntie+1;
yy(i)=ry(i);
ry(i)=0;
end
for i=1:n
if (ry(i)~=0) continue; end
yy(i)=yy(i)+ntie*0.5;
ry(i)=n+1;
end
end
rs=0;
rs=sum((xx-yy).^2);
spearm=1-((6*rs)/(n*(n^2-1)));
```

2.2 Statistic test for between-image difference

Before statistic test for between-image difference, matrices A and B are transformed to vectors A_1 and B_1

respectively, i.e., (A_1, B_1)={(a_i, b_i) | i=1, 2, ..., p} (Zhang, 2018b). A statistically significant difference between two images means the remarkable difference in structure and / or grayscale of two images.

2.2.1 Signed test of matched samples for between-image difference

We assume the difference between the matched samples in the vectors A_1 and B_1 comes from a distribution whose median is zero. The differences A_1-B_1 are assumed to come from an arbitrary continuous distribution.

Full Matlab codes are as follows:

```
alpha=input('Input significance level (e.g., 0.01): ');
RGB1=imread('geneExpression.png'); RGB2=imread('soluableProteins.png');
Gray1=rgb2gray(RGB1); Gray2=rgb2gray(RGB2);
m=round((size(Gray1,1)+size(Gray2,1))/2); n=round((size(Gray1,2)+size(Gray2,2))/2);
I=imresize(Gray1,[m n]); J=imresize(Gray2,[m n]);
A=im2double(I); B=im2double(J);
A1=reshape(A, m*n, 1);    %Transform matrix A to vector A1
B1=reshape(B, m*n, 1);    %Transform matrix B to vector B1
p=signtest(A1,B1,'alpha',alpha);
if (p<=alpha) fprintf(['Difference is statistically significant (p=' num2str(p) ')\n']);
else fprintf(['Difference is not statistically significant (p=' num2str(p) ')\n']); end
```

2.2.2 Wilcoxon signed rank test for zero median of between-image difference

We assume the difference between the matched samples in the vectors A_1 and B_1 comes from a distribution whose median is zero. The differences A_1-B_1 are assumed to come from a continuous distribution, symmetric about its median. Full Matlab codes are as follows:

```
alpha=input('Input significance level (e.g., 0.01): ');
RGB1=imread('geneExpression.png'); RGB2=imread('soluableProteins.png');
Gray1=rgb2gray(RGB1); Gray2=rgb2gray(RGB2);
m=round((size(Gray1,1)+size(Gray2,1))/2); n=round((size(Gray1,2)+size(Gray2,2))/2);
I=imresize(Gray1,[m n]); J=imresize(Gray2,[m n]);
A=im2double(I); B=im2double(J);
A1=reshape(A, m*n, 1);    %Transform matrix A to vector A1
B1=reshape(B, m*n, 1);    %Transform matrix B to vector B1
p=signrank(A1,B1,'alpha',alpha);
if (p<=alpha) fprintf(['Difference is statistically significant (p=' num2str(p) ')\n']);
else fprintf(['Difference is not statistically significant (p=' num2str(p) ')\n']); end
```

2.2.3 *t*-test for between-image difference

Assume the difference A_1-B_1 between two vectors A_1 and B_1 follows the normal distribution. If there is not statistic difference between images A and B, the mean of A_1-B_1 should be zero. The *t*-test can be used to test the statistic difference (Zhang, 2018). Full Matlab codes are as follows:

```
alpha=input('Input significance level (e.g., 0.01): ');
RGB1=imread('geneExpression.png'); RGB2=imread('soluableProteins.png');
Gray1=rgb2gray(RGB1); Gray2=rgb2gray(RGB2);
m=round((size(Gray1,1)+size(Gray2,1))/2); n=round((size(Gray1,2)+size(Gray2,2))/2);
```

```
I=imresize(Gray1,[m n]); J=imresize(Gray2,[m n]);
A=im2double(I); B=im2double(J);
A1=reshape(A, m*n, 1);   %Transform matrix A to vector A1
B1=reshape(B, m*n, 1);   %Transform matrix B to vector B1
[h,p,ci,stats]=ttest(A1-B1,0,alpha,0);
if (h==1) fprintf(['Difference is statistically significant (p=' num2str(p) ')\n']);
else fprintf(['Difference is not statistically significant (p=' num2str(p) ')\n']); end
```

2.3 Similarity index

Similarity index is defined as

$$S=1-\Sigma\Sigma|a_{ij}-b_{ij}|/(m\times n)$$

where $1\leq S\leq 1$. A greater S means the greater similarity in structure and / or grayscale between two images.

Full Matlab codes are as follows:

```
RGB1=imread('geneExpression.png'); RGB2=imread('soluableProteins.png');
Gray1=rgb2gray(RGB1); Gray2=rgb2gray(RGB2);
m=round((size(Gray1,1)+size(Gray2,1))/2); n=round((size(Gray1,2)+size(Gray2,2))/2);
I=imresize(Gray1,[m n]); J=imresize(Gray2,[m n]);
A=im2double(I); B=im2double(J);
S=1-sum(sum(abs(A-B)))/(m*n);
fprintf(['Similarity index S=' num2str(S) '\n']);
```

The executable GUI software (see supplementary material), imageAlignment, of the methods above is partly indicated in Fig. 1.

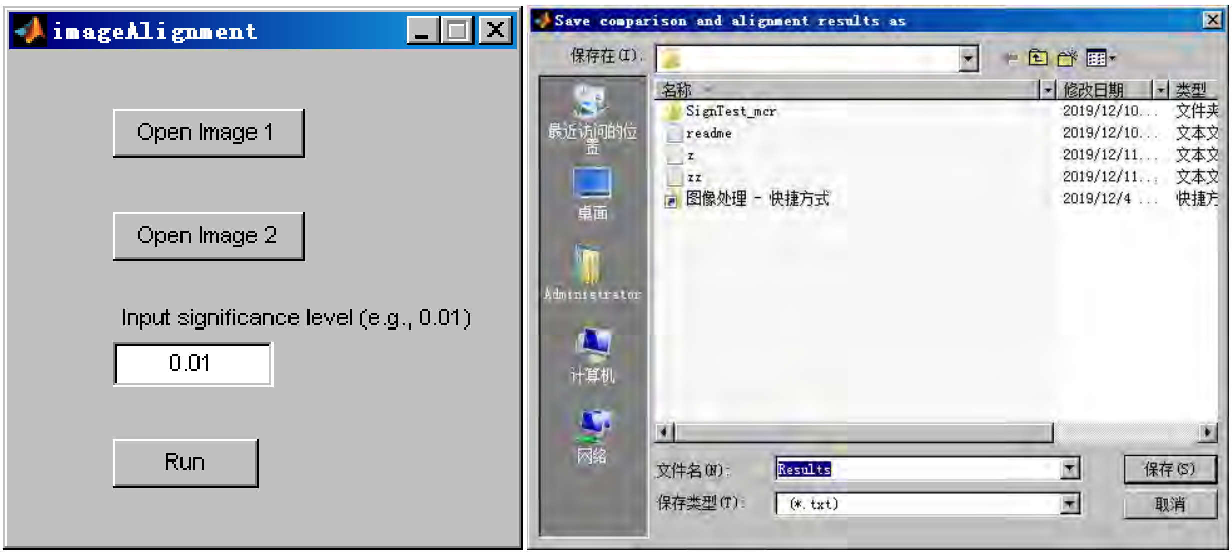


Fig. 1 The executable GUI software of the methods.

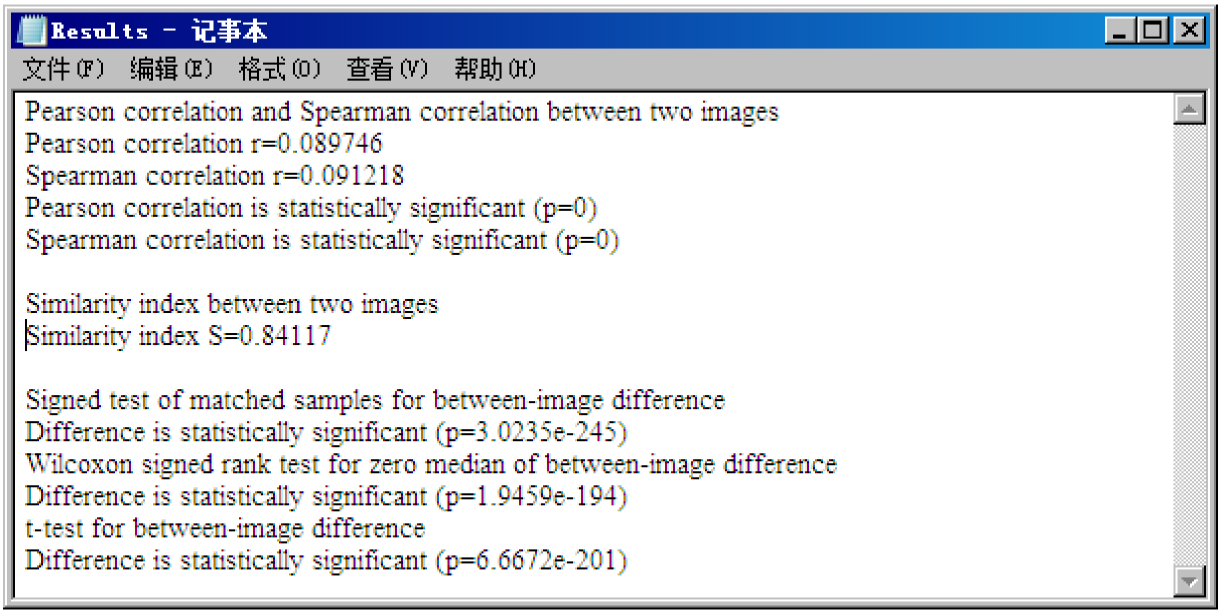
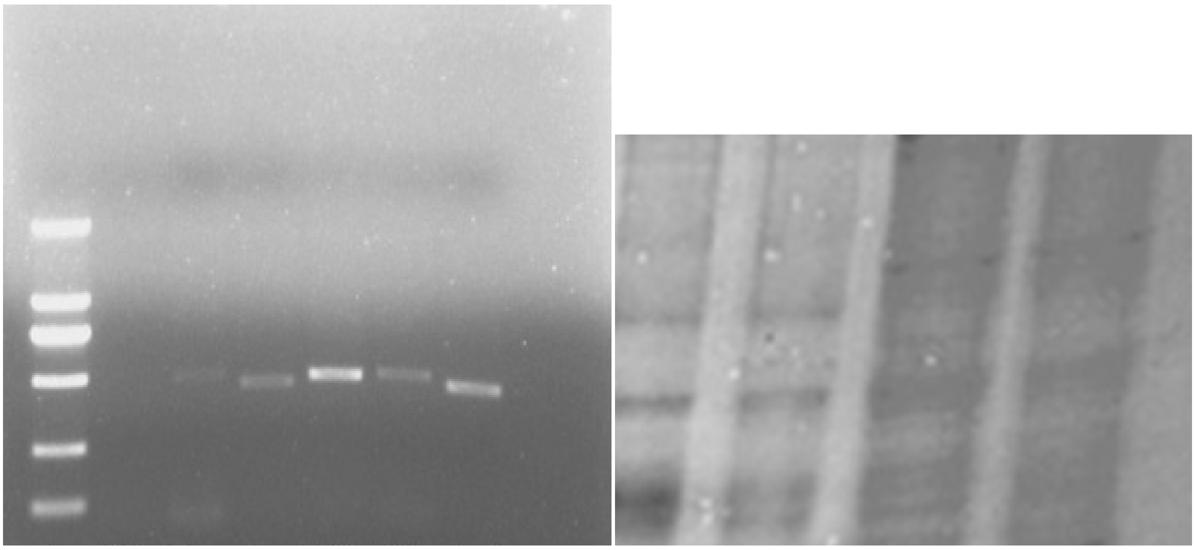


Fig. 2 The saved results.

3 Application example

Two images (supplementary material), indicated in Fig. 3, are used for analysis. In present study, the statistic significance level is set to be 0.01. Using the methods above, the results are as follows (Fig. 2).



(a) (b)

Fig. 3 Two biochemical images. (a) geneExpresion.png; (b) soluableProteins.png.

3.1 Between-image Pearson correlation and Spearman correlation

(1) Two different images

Between geneExpression.png and soluableProteins.png:

Pearson correlation r=0.089746. Pearson correlation is statistically significant (p=0).

Spearman correlation r=0.091218. Pearson correlation is statistically significant (p=0)
(2) Same images
Between geneExpression.png and geneExpression.png:
Pearson correlation r=1. Pearson correlation is not statistically significant.
Spearman correlation r=1. Pearson correlation is not statistically significant.

3.2 Statistic test for between-image difference

(1) Two different images
Between geneExpression.png and soluableProteins.png:
Signed test: difference is statistically significant (p=3.0235e-245).
Wilcoxon signed rank test for zero median: difference is statistically significant (p=1.9459e-194).
t-test: difference is statistically significant (p=6.6672e-201).
(2) Same images
Between geneExpression.png and geneExpression.png:
Signed test: difference is not statistically significant (p=1).
Wilcoxon signed rank test for zero median: difference is not statistically significant (p=1).
t-test: difference is not statistically significant (p=NaN).

3.3 Similarity index

(1) Two different images
Between geneExpression.png and soluableProteins.png: Similarity index S=0.84117.
(1) Same images
Between geneExpression.png and geneExpression.png: Similarity index S=1.

Acknowledgment

We are thankful to the support of The National Key Research and Development Program of China (2017YFD0201204), and Discovery and Crucial Node Analysis of Important Biological and Social Networks (2015.6-2020.6), from Yangling Institute of Modern Agricultural Standardization.

References

Zhang WJ. 2012. Computational Ecology: Graphs, Networks and Agent-based Modeling. World Scientific, Singapore

Zhang WJ. 2015. Calculation and statistic test of partial correlation of general correlation measures. Selforganizology, 2(4): 65-77

Zhang WJ. 2016. Network pharmacology: A further description. Network Pharmacology, 1(1): 1-14

Zhang WJ. 2018a. Fundamentals of Network Biology. World Scientific Europe, London, UK

Zhang WJ. 2018b. Network matrix based methods for between-network comparison. Network Biology, 8(4): 144-152

Zhang WJ, Li X. 2015. General correlation and partial correlation analysis in finding interactions: with Spearman rank correlation and proportion correlation as correlation measures. Network Biology, 5(4): 163-168